Review in Probability

Contents

1 Review

1.1 Continuous Random Variables

A **continuous random variable** is one where its cumulative distribution function is continuous.
Let X be a continuous random variable, and let $F_X(x) = P(X \leq x)$ be its cumulative distribution function. Then

$$\begin{aligned} P(a < X \leq b) &= P(X \leq b) - P(X \leq a) \\ &= F_X(b) - F_X(a) \\ &= \int_a^b \frac{d}{dx}[F_X(x)]\, dx \quad \text{(Fundamental Theorem of Calculus).} \end{aligned}$$

We let $f_X(x) = \frac{d}{dx}[F_X(x)]$. The function $f_X(x)$ is called the **probability density function**.
For a set $E \subset \mathbb{R}$, to compute $P(X \in E)$, we have the integral

$$\boxed{P(X \in E) = \int_E f_X(x)\, dx.}$$

Example 1. Let X be a continuous random variable with probability density function

$$f(x) = \begin{cases} 2x & 0 < x < 1, \\ 0 & \text{otherwise.} \end{cases}$$

Then the cumulative distribution function of X is given by

$$F(x) = \begin{cases} 1 & x > 1, \\ x^2 & 0 < x \leq 1, \\ 0 & x \leq 0. \end{cases}$$

For a continuous random variable X, the **expected value** is

$$E[X] = \int_{-\infty}^{\infty} x f_X(x)\, dx.$$

For a function $g(X)$, we have

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

The **variance** is

$$\begin{aligned} \mathrm{Var}(X) = E[(X - E[X])^2] &= \int_{-\infty}^{\infty} (x - E[X])^2 f_X(x)\, dx \\ &= E[X^2] - E[X]^2. \end{aligned}$$

1.1.1 Uniform Random Variable

A **uniform random variable** over $[a, b]$ is a random variable with probability density function

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

Example 2. Suppose a person shows up at a train station uniformly between 12pm and 1pm. The train leaves at 12:45pm. What is the probability that the person misses the train?
Solution: Intuitively, the person misses the train if he arrives after 12:45, there are 15 minutes where if he arrives, he misses the train out of 60 minutes, so the answer is $\frac{15}{60} = \frac{1}{4}$. Using the uniform random variable, let X be uniform over $[0, 60]$. Then the person misses the train when $X > 45$. Thus,

$$P(X > 45) = \int_{45}^{60} \frac{1}{60} dx = \frac{15}{60}.$$

1.1.2 Normal Random Variable

A **normal random variable** with mean μ and standard deviation σ is a random variable with probability density function

$$\boxed{f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} \qquad -\infty < x < \infty.}$$

Given any random variable, we can subtract the mean and divide by the standard deviation to get a new random variable with mean 0 and standard deviation 1. That is, let X be any random variable with mean μ and standard deviation σ. Then $Z = \frac{X-\mu}{\sigma}$ is a random variable with mean 0 and standard deviation 1. The cdf $F_Z(x)$ of the standard normal is usually denoted as $\Phi(x)$. Using this technique, we can use a Z-table to compute probabilities. Note that the Z-table only has numbers starting from 0, so for negative numbers, we need to do some algebra. One of the problems addresses this.
For example, let X be normal with mean μ and standard deviation σ. Then to find $P(a < X < b)$, we have

$$\begin{aligned} P(a < X < b) &= P\left(\frac{a-\mu}{\sigma} < \frac{X-\mu}{\sigma} < \frac{b-\mu}{\sigma}\right) \\ &= P\left(\frac{a-\mu}{\sigma} < Z < \frac{b-\mu}{\sigma}\right) \\ &= \Phi\left(\frac{b-\mu}{\sigma}\right) - \Phi\left(\frac{a-\mu}{\sigma}\right). \end{aligned}$$

To finish it off, use the Z-table to find the value of $\Phi(\frac{b-\mu}{\sigma})$ and subtract off $\Phi(\frac{a-\mu}{\sigma})$.

Example 3. Annual snowfall at a specific location is modeled by a normal random variable with mean $\mu = 60$ inches, and standard deviation $\sigma = 20$. What is the probability this year's snowfall will be between 40 and 80 inches? What is the probability that it will exceed 80 inches?
Solution: Let X be a normal random variable with mean 60 and standard deviation 20. Then we want to find $P(40 < X < 80)$. We have

$$\begin{aligned} P(40 < X < 80) &= P\left(\frac{40-60}{20} < \frac{X-60}{20} < \frac{80-60}{20}\right) \\ &= P(-1 < Z < 1) \\ &= \Phi(1) - \Phi(-1) \\ &= 0.8413 - 0.1587 = 0.6826. \end{aligned}$$

For the second part, we want $P(X > 80)$, so

$$\begin{aligned} P(X > 80) &= P\left(\frac{X-60}{20} > \frac{80-60}{20}\right) \\ &= P(Z > 1) \\ &= 1 - P(Z < 1) \\ &= 1 - \Phi(1) \\ &\approx 1 - 0.8413 = 0.1587. \end{aligned}$$

1.1.3 Binomial Approximation

Using the normal random variable, we can approximate the binomial random variable with parameters n and p. This is called the **DeMoivre-Laplace approximation**. The mean of such a binomial is np and the variance is $np(1-p)$. Let X be a normal random variable with mean np and variance $np(1-p)$. However, the binomial is discrete and normal is continuous, so we need to apply a **continuity correction**.

1. For $P(X = n)$ use $P(n - \frac{1}{2} < X < n + \frac{1}{2})$,
2. For $P(X > n)$ use $P(X > n + \frac{1}{2})$,
3. For $P(X \le n)$ use $P(X < n + \frac{1}{2})$,
4. For $P(X < n)$ use $P(X < n - \frac{1}{2})$,
5. For $P(X \ge n)$ use $P(X > n - \frac{1}{2})$.

Example 4. I flip a fair coin 100 times. Approximately what is the probability that I get between 55 and 60 heads inclusively?
Solution: Let X be a binomial random variable, with $n = 100$, and $p = \frac{1}{2}$. The mean of X is 50. The variance is 25. Then

$$\begin{aligned}
P(55 \le X \le 60) &= P(54.5 < X < 60.5) && \text{(continuity correction)} \\
&= P\left(\frac{54.5-50}{\sqrt{25}} \le \frac{X-50}{\sqrt{25}} \le \frac{60.5-50}{\sqrt{25}}\right) \\
&= P(0.9 < Z < 2.1) && \left(\text{replace } \frac{X-\mu}{\sigma} \text{ with } Z\right) \\
&= \Phi(2.1) - \Phi(0.9) \\
&\approx 0.9821 - 0.8159 = 0.1662.
\end{aligned}$$

1.1.4 Exponential Random Variable

The **exponential random variable** with mean $1/\lambda$ has the probability density function

$$\boxed{f(x) = \begin{cases} \lambda e^{-\lambda x} & x > 0 \\ 0 & \text{otherwise.} \end{cases}}$$

Exponential random variables are used to model time until some event happens. One property of the exponential random variable is the **memoryless property**, which is

$$P(X > s + t | X > s) = P(X > t), \qquad P(X < s + t | X > s) = P(X < t).$$

This property states that if we know time passed and an event did not happen, then we can just start counting time again from zero.

1.1.5 Gamma Random Variable

The gamma function can be viewed as an extension of the factorial function to all positive numbers. This is given by

$$\Gamma(s) = \int_0^\infty x^{s-1}e^{-x}\,dx.$$

We have the following properties of the gamma function

$$\Gamma(n) = (n-1)!, \qquad \Gamma(s) = (s-1)\Gamma(s-1), \qquad \Gamma\left(\frac{1}{2}\right) = \sqrt{\pi}.$$

Based on this information, we can define a random variable, whose pdf is similar to the integrand of the gamma function. Thus, we define the **gamma random variable**, with parameters α, λ to have probability density function

$$\boxed{f(x) = \begin{cases} \dfrac{\lambda^\alpha}{\Gamma(\alpha)} x^{\alpha-1}e^{-\lambda x} & x > 0 \\ 0 & \text{otherwise.} \end{cases}}$$

1.1.6 Cauchy Random Variable

The **Cauchy random variable** is given by the pdf

$$\boxed{f(x) = \frac{1}{\pi(1+x^2)} \qquad -\infty < x < \infty.}$$

One important property of the Cauchy random variable is that its expected value cannot be defined.

1.1.7 Transformations of a Random Variable

For a discrete random variable X, let Y be another random variable defined to be $Y = g(X)$. In the case where $g(X)$ is an injective function, we have $P(Y = y) = P(g(X) = y) = P(X = g^{-1}(y))$. But if $g(X)$ is not injective, then we should split up the cases into injective pieces.
For a continuous random variable X, let $Y = g(X)$. Then the cdf of Y is $P(Y \le y) = P(g(X) \le y) = P(X \le g^{-1}(y)) = \int_{-\infty}^{g^{-1}(y)} f_X(x)\,dx$. Then we differentiate with respect to y to get the pdf of Y, which, by Fundamental Theorem of Calculus and the chain rule, is

$$\boxed{f_Y(y) = f_X(g^{-1}(y))\frac{d}{dy}[g^{-1}(y)].}$$

Though this is the formula, we have to be careful in computations, especially in finding the bounds. Again, the formula does not work if $g(X)$ is not an injective function, and we would need to split up the function into injective pieces.

Example 5. I flip a fair coin until I get heads. Let X be the number of flips it takes for me to get heads. Let Y be the number of tails in this situation. What is the probability mass function of Y?
Solution: The pmf of X is $P(X = n) = \frac{1}{2}^n$. In order for $X = n$, we need to have $n-1$ tails, which means that $Y = X - 1$. Then we have that $P(Y = k) = P(X - 1 = k) = P(X = k+1) = (\frac{1}{2})^{k+1}$. For a complete answer, the pmf is

$$P(Y = k) = \begin{cases} (\frac{1}{2})^{k+1} & y = 0, 1, 2, \ldots \\ 0 & \text{otherwise.} \end{cases}$$

Example 6. Let X be uniform over $[0,2]$. Let $Y = 2X - 1$. What is the pdf of Y? What is the distribution of Y?
Solution: We know X has pdf $f_X(x) = \frac{1}{2}, 0 < x < 2$ and 0 otherwise. Then we have that

$$\begin{aligned} 1 &= \int_0^2 \frac{1}{2} dx \\ &= \int_{-1}^3 \frac{1}{2}\frac{dy}{2} \quad \text{substitute } y = 2x-1, dy = 2dx \\ &= \int_{-1}^3 \frac{1}{4} dy. \end{aligned}$$

Then the pdf of Y is

$$f_Y(y) = \begin{cases} \frac{1}{4} & -1 < y < 4 \\ 0 & \text{otherwise.} \end{cases}$$

Therefore, the distribution of Y is a uniform distribution over $[-1,3]$.

Example 7. Let X be normal with mean 0 and variance 1. Let $Y = X^2$. What is the pdf of Y? What is the distribution of Y?
Solution: We present two solutions.

1. We know X has pdf $f_X(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}$. Then we have that

$$\begin{aligned} 1 &= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}\, dx \\ &= \int_{-\infty}^{0} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}\, dx + \int_{0}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}\, dx \end{aligned}$$

   Now we consider the two integrals separately. First, we have that for $x > 0$,

$$\begin{aligned} \int_0^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}\, dx &= \int_0^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}y}\, \frac{dy}{2\sqrt{y}} \quad \text{substitute } \sqrt{y} = x, \frac{1}{2\sqrt{y}}dy = dx \\ &= \int_0^{\infty} \frac{1}{\sqrt{2\pi}} \frac{1}{2\sqrt{y}} e^{-\frac{1}{2}y}\, dy. \end{aligned}$$

   For $x < 0$, we have

$$\begin{aligned} \int_{-\infty}^{0} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}\, dx &= \int_{\infty}^{0} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}y}\, \frac{dy}{-2\sqrt{y}} \quad \text{substitute } -\sqrt{y} = x, -\frac{1}{2\sqrt{y}}dy = dx \\ &= \int_0^{\infty} \frac{1}{\sqrt{2\pi}} \frac{1}{2\sqrt{y}} e^{-\frac{1}{2}y}\, dy. \end{aligned}$$

   Taking the sum, we have

$$\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}\, dx = 2\int_0^{\infty} \frac{1}{\sqrt{2\pi}} \frac{1}{2\sqrt{y}} e^{-\frac{1}{2}y}\, dy.$$

   Therefore, the pdf of Y is

$$f_Y(y) = \begin{cases} \frac{1}{\sqrt{2\pi}} \frac{1}{\sqrt{y}} e^{-\frac{1}{2}y} & 0 < y < \infty \\ 0 & \text{otherwise.} \end{cases}$$

2. Consider $F_Y(y) = P(Y \leq y)$. Then

$$\begin{aligned} P(Y \leq y) &= P(X^2 \leq y) \\ &= P(-\sqrt{y} \leq X \leq \sqrt{y}) \\ &= P(X \leq \sqrt{y}) - P(X < -\sqrt{y}) \\ &= F_X(\sqrt{y}) - F_X(-\sqrt{y}). \end{aligned}$$

Then we differentiate $F_Y(y)$ to get the pdf $f_Y(y)$ of Y. By applying the chain rule to the fundamental theorem of calculus, we get

$$\begin{aligned} f_Y(y) &= \frac{d}{dy}\left[\int_{-\infty}^{\sqrt{y}} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}\, dx - \int_{-\infty}^{-\sqrt{y}} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}\, dx\right] \\ &= \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}y}\left(\frac{1}{2\sqrt{y}}\right) - \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}y}\left(-\frac{1}{2\sqrt{y}}\right) \\ &= \frac{1}{\sqrt{2\pi}} \frac{1}{\sqrt{y}} e^{-\frac{1}{2}y}. \end{aligned}$$

Since X is between $-\infty$ and ∞, we know that Y is between 0 and ∞. Putting that together, the pdf of Y is

$$f_Y(y) = \begin{cases} \frac{1}{\sqrt{2\pi}} \frac{1}{\sqrt{y}} e^{-\frac{1}{2}y} & 0 < y < \infty \\ 0 & \text{otherwise.} \end{cases}$$

The distribution of Y is known as the **chi-squared distribution** with 1 degree of freedom.

1.2 Jointly Distributed Random Variables

Given two random variables X, Y, we can talk about their probabilities at the same time. This is known as **jointly distributed random variables**. The joint cdf is a multivariable function given by

$$F_{X,Y}(x, y) = P(\{X \leq x\} \cap \{Y \leq y\}).$$

Taking partial derivatives, we get that the joint pdf is

$$\boxed{f_{X,Y}(x, y) = \frac{\partial^2}{\partial x \partial y} F_{X,Y}(x, y).}$$

To compute probabilities, let $E \in \mathbb{R}^2$. Then

$$\boxed{P((X, Y) \in E) = \iint_E f_{X,Y}(x, y)\, dA.}$$

As an example, suppose we wanted to find the probability that $X < Y$. Then we have

$$P(X < Y) = \iint_{\{x<y\}} f_{X,Y}(x, y)\, dA = \int_{-\infty}^{\infty} \int_{-\infty}^{y} f(x, y) dx dy.$$

Example 8. Let X and Y have joint pdf

$$f(x, y) = \begin{cases} 2x & 0 < x < 1, 0 < y < 1 \\ 0 & \text{otherwise.} \end{cases}$$

Then

$$P(X < Y) = \int_0^1 \int_0^y 2x dx dy = \int_0^1 \int_x^1 2x dy dx = \frac{1}{3}.$$

Example 9. Let X and Y have joint pdf

$$f(x,y) = \begin{cases} 6x & 0 < x < 1, 0 < y < 1, 0 < x + y < 1 \\ 0 & \text{otherwise.} \end{cases}$$

Then

$$P(X < Y) = \int_0^{\frac{1}{2}} \int_x^{1-x} 2x dy dx = \frac{1}{4}.$$

In this case, it is better to do $dydx$.

Exercise: Convince yourself that the bounds above are correct. Attempt to find $P(X < Y)$ through a double integral that integrates in the order $dxdy$.

1.2.1 Marginal Distributions

Given a joint cdf of X and Y, we note that X and Y are each random variables themselves. So we have that X has its own pdf, and so does Y. These are called the **marginal density functions**, and they are given by

$$\boxed{f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)\, dy, \qquad f_Y(y) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)\, dx.}$$

To remember which variable we integrate, the marginal density of X is a function of x, so we must integrate out y. Similarly, the marginal density of Y is a function of y, so we integrate out x.

Example 10. Let X and Y be random variables with joint density function

$$f(x,y) = \begin{cases} 2e^{-(x+2y)} & 0 < x < \infty, 0 < y < \infty, \\ 0 & \text{otherwise.} \end{cases}$$

We will find the marginal densities of X and Y.
First we note that $1 = \int_0^\infty \int_0^\infty 2e^{-(x+2y)} dydx = \int_0^\infty \int_0^\infty 2e^{-(x+2y)} dxdy$. To get the marginal density of X, we integrate in the order $dydx$ and then drop the outer most integral, that is, the dx integral. To get the marginal density of Y, we integrate in the order $dxdy$ and then drop the dx integral. Thus, the marginal density of X is $f_X(x) = \int_0^\infty 2e^{-(x+2y)} dy, 0 < x < \infty$ and the marginal density of Y is $f_Y(y) = \int_0^\infty 2e^{-(x+2y)} dx, 0 < y < \infty$. In conclusion, we have

$$f_X(x) = \begin{cases} e^{-x} & 0 < x < \infty, \\ 0 & \text{otherwise,} \end{cases} \qquad f_Y(y) = \begin{cases} 2e^{-2y} & 0 < y < \infty, \\ 0 & \text{otherwise.} \end{cases}$$

Example 11. Let X and Y be random variables with joint density function

$$f(x,y) = \begin{cases} 6e^{-(x+2y)} & 0 < x < y < \infty, \\ 0 & \text{otherwise.} \end{cases}$$

In this case, $1 = \int_0^\infty \int_x^\infty 6e^{-(x+2y)} dydx = \int_0^\infty \int_0^y 2e^{-(x+2y)} dxdy$. So the marginal density of X is $f_X(x) = \int_x^\infty 6e^{-(x+2y)} dy, 0 < x < \infty$ and the marginal density of Y is $f_Y(y) = \int_0^y 6e^{-(x+2y)} dx, 0 < y < \infty$. Therefore, we get

$$f_X(x) = \begin{cases} \frac{1}{2}e^{-3x} & 0 < x < \infty, \\ 0 & \text{otherwise,} \end{cases} \qquad f_Y(y) = \begin{cases} e^{-2y}(1 - e^{-y}) & 0 < y < \infty, \\ 0 & \text{otherwise.} \end{cases}$$

1.2.2 Independence of Random Variables

Recall that two events E, F are independent if $P(E \cap F) = P(E)P(F)$. Using this, we get that two random variables X, Y are **independent** if and only if $F_{X,Y}(x, y) = F_X(x)F_y(y)$. Equivalently, X, Y are independent if and only if $f_{X,Y}(x, y) = f_X(x)f_Y(y)$.

Example 12. Let X and Y be random variables with joint density function

$$f(x, y) = \begin{cases} 2e^{-(x+2y)} & 0 < x < \infty, 0 < y < \infty, \\ 0 & \text{otherwise.} \end{cases}$$

We have X and Y are independent, since after computing the marginal densities, we get

$$f(x, y) = 2e^{-(x+2y)} = e^{-x}(2e^{-2y}) = f_X(y)f_Y(y).$$

The phrase **independent and identically distributed** means that a collection of random variables $X_1, X_2, \ldots, X_n$ are independent and have the same probability density function, call it $f(x)$. Therefore, the joint density function of $X_1, X_2, \ldots, X_n$ is

$$f(x_1, \ldots, x_n) = f(x_1) \times f(x_2) \times \cdots \times f(x_n).$$

1.2.3 Sum of Independent Random Variable

The **sum of independent random variables** has a formula. Let X and Y be independent random variables. Then the pdf of $X + Y$ is

$$\boxed{f_{X+Y}(a) = \int_{-\infty}^{\infty} f_X(a - y)f_Y(y)\, dy.}$$

This kind of integral can be seen in other fields of math and is called *convolution*. The example that everyone does is the sum of two uniform random variables. Computation is in the appendix.
Now we look at special cases of sums of random variables.

1. **Sum of Poisson** Let X be a Poisson random variable with parameter λ_1 and Y be a Poisson random variable with parameter λ_2. Then $X + Y$ is Poisson as well, with parameter $\lambda_1 + \lambda_2$.

2. **Sum of Normal** Now let X be a normal random variable with mean μ_1 and variance σ_1^2 and Y be a normal random variable with mean μ_2 and variance σ_2^2. Then $X + Y$ is normal with mean $\mu_1 + \mu_2$ and variance $\sigma_1^2 + \sigma_2^2$. Note that the standard deviation of the sum of two normals is not the sum of the standard deviations.

3. **Sum of Exponential** The last case is of the exponential random variable. Let X and Y be exponential with mean $1/\lambda$. Then $X + Y$ is a gamma random variable, with parameters 2 and λ, that is, the pdf of $X + Y$ is
$$f_{X+Y}(x) = \frac{\lambda^2}{\Gamma(2)} xe^{-\lambda x}.$$

For all three special cases above, the easiest way to find the sum of independent random variables is through moment-generating functions, seen later.

1.2.4 Order Statistics

Let $X_1, X_2, \ldots, X_n$ be independent and identically distributed. The **order statistics** are $Y_1, \ldots, Y_n$, where Y_1 is the smallest value of $X_1, \ldots, X_n$, Y_2 is the second smallest value, etc. The joint pdf of the order statistics is

$$\boxed{f(y_1, \ldots, y_n) = n! \prod_{i=1}^{\infty} f(y_i).}$$

The marginal density of Y_k is

$$\boxed{f_{Y_k}(y_k) = \binom{n}{k-1, 1, n-k} F(y_k)^{k-1} f(y_k)[1 - F(y_k)]^{n-k}.}$$

See appendix for more information.

1.2.5 Conditional Distributions

Recall the definition of conditional probability to be

$$P(E|F) = \frac{P(E \cap F)}{P(F)}.$$

When X and Y are discrete, using this definition, we define the **conditional mass function** of X given $Y = y$ to be

$$P(X = x|Y = y) = \frac{P(X = x, Y = y)}{P(Y = y)}.$$

Note that this function is a function of x, and y is usually a fixed value. When X and Y are continuous, the formula for the **conditional density function** becomes

$$f_{X|Y}(x|y) = \frac{f_{X,Y}(x, y)}{f_Y(y)}.$$

This does not directly follow from the definition of conditional probability, so we need more explanation on where it comes from. See appendix for details.

1.2.6 Transformations of Joint Random Variables

Let X, Y be random variables, let $U = u(X, Y)$ and $V = v(X, Y)$. Then U and V have a joint distribution. In order to find it, first, we need to solve for X and Y in terms of U and V. That is, we find functions such that we are able to write $X = x(U, V)$, and $Y = y(U, V)$. Then we have the formula,

$$f_{U,V}(u, v) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \frac{f(x(u, v), y(u, v))}{\left|\frac{\partial(u,v)}{\partial(x,y)}\right|} dudv,$$

where

$$\left|\frac{\partial(u, v)}{\partial(x, y)}\right| = \det \begin{pmatrix} \frac{\partial u}{\partial x} & \frac{\partial u}{\partial y} \\ \frac{\partial v}{\partial x} & \frac{\partial v}{\partial y} \end{pmatrix}.$$

Therefore, the joint density function of U and V is

$$\frac{f(x(u, v), y(u, v))}{\left|\frac{\partial(u,v)}{\partial(x,y)}\right|}.$$

1.3 Properties of Expectation

1.3.1 Linearity of Expectation

Linearity of expectation is a strong property. To use it in problems, let X be a random variable, and we want to find $E[X]$, but it might be difficult to do directly. If we are able to write $X = X_1 + X_2 + \cdots + X_n$, such that finding the expectations of X_is are easy, then we have that

$$E[X] = E[X_1] + E[X_2] + \cdots + E[X_n].$$

Example 13. I flip a fair coin 5 times. Let X be the number of times two consecutive heads shows up. For example, if I get $HHTHT$, then $X = 1$, but if I get $HHHTH$, then $X = 2$. Find $E[X]$.
Solution: Define the random variable X_1 by

$$X_1 = \begin{cases} 1 & \text{if the first and second flips land heads} \\ 0 & \text{otherwise.} \end{cases}$$

Similarly, define X_2, X_3, X_4 by $X_2 = 1$ if the second and third flips land heads, $X_3 = 1$ if the third and fourth flips land heads, $X_4 = 1$ if the fourth and fifth flips land heads.
Then $X = X_1 + X_2 + X_3 + X_4$. For any i, we have that $E[X_i] = \frac{1}{4}$. (Why?) Therefore,

$$\begin{aligned} E[X] &= E[X_1 + X_2 + X_3 + X_4] \\ &= E[X_1] + E[X_2] + E[X_3] + E[X_4] && \text{(linearity of expectation)} \\ &= \frac{1}{4} + \frac{1}{4} + \frac{1}{4} + \frac{1}{4} = 1. \end{aligned}$$

1.3.2 Covariance and Correlation

Let X and Y be random variables. Then we would like to measure the relationship between the two, that is, a measure of how dependent X and Y are. This is done by **covariance** of X and Y, defined by

$$\text{Cov}(X, Y) = E[(X - E[X])(Y - E[Y])].$$

Using some algebra and linearity of expectation, we get that the covariance is equal to

$$\boxed{\text{Cov(X, Y)} = E[XY] - E[X]E[Y].}$$

Covariance is positive when X large makes Y more likely to be large, and covariance is negative when X is large makes Y more likely to be small.

Properties of Independent Random Variables

If X and Y are independent, then

1. $E[XY] = E[X]E[Y]$,
2. $\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y)$,
3. $\text{Cov}(X, Y) = 0$.

However, if any of the above 3 are true, then it does not imply independence. In other words, we cannot use any of the three facts above to show that X and Y are independent.

Properties of Covariance

1. $\text{Cov}(X, Y) = \text{Cov}(Y, X)$,
2. $\text{Cov}(aX, bY) = ab\,\text{Cov}(X, Y)$,
3. $\text{Cov}(X + Y, Z) = \text{Cov}(X, Z) + \text{Cov}(Y, Z)$,
4. $\text{Cov}(X, Y + Z) = \text{Cov}(X, Y) + \text{Cov}(X, Z)$.

Example 14. We have

$$\operatorname{Var}(X+Y) = \operatorname{Var}(X) + \operatorname{Var}(Y) + 2\operatorname{Cov}(X,Y).$$

To see this, we have

$$\begin{aligned}
\operatorname{Var}(X+Y) &= E[(X+Y)^2] - (E[X+Y])^2 \\
&= E[X^2 + Y^2 + 2XY] - (E[X] + E[Y])^2 \\
&= E[X^2] + E[Y^2] + 2E[XY] - E[X]^2 - E[Y]^2 - 2E[X]E[Y] \\
&= \left[E[X^2] - E[X]^2\right] + \left[E[Y^2] - E[Y]^2\right] + 2\left[E[XY] - E[X]E[Y]\right] \\
&= \operatorname{Var}(X) + \operatorname{Var}(Y) + 2\operatorname{Cov}(X,Y).
\end{aligned}$$

Correlation

Covariance ideally measures the relationship between two random variables. Covariance has an issue where it depends on the units of the random variable. For example, we would like $\operatorname{Cov}(X,Y)$ to be equal to $\operatorname{Cov}(2X,2Y)$, as the relationship between X and Y did not change, but in reality, $\operatorname{Cov}(2X,2Y) = 4\operatorname{Cov}(X,Y)$.
The **correlation coefficient** adjusts for the differences in the variances of X and Y, and is defined to be

$$\boxed{\rho(X,Y) = \frac{\operatorname{Cov}(X,Y)}{\sqrt{\operatorname{Var}(X)\operatorname{Var}(Y)}}.}$$

The correlation coefficient is always between -1 and 1, that is,

$$\boxed{-1 \leq \rho(X,Y) \leq 1.}$$

1.3.3 Conditional Expectation

Let X and Y be random variables. Recall that the conditional density of X given $Y = y$ is a function of x. Using the usual definition of expectation, the **conditional expectation** is

$$\boxed{E[X|Y=y] = \int_{-\infty}^{\infty} x f_{X|Y}(x|y)\,dx = \int_{-\infty}^{\infty} x \frac{f_{X,Y}(x,y)}{f_Y(y)}\,dx.}$$

For a function $g(X)$, we have

$$\boxed{E[g(X)|Y=y] = \int_{-\infty}^{\infty} g(x) f_{X|Y}(x|y)\,dx.}$$

Let X be any random variable, and let Y be discrete. The **law of total expectation** is

$$E[X] = \sum_y E[X|Y=y]P(Y=y).$$

In the case that Y is continuous, the law of total expectation becomes

$$E[X] = \int_{-\infty}^{\infty} E[X|Y=y] f_Y(y)\,dy.$$

1.3.4 Moments and Moment Generating Functions

Let X be a (continous) random variable. The nth **moment** is defined to be $E[X^n] = \int_{-\infty}^{\infty} x^n f(x)\,dx$. The **moment generating function** of X, denoted by $M_X(t)$, is

$$E[e^{tX}] = \int_{-\infty}^{\infty} e^{tx} f(x)\,dx.$$

Example 15. Let X be Bernoulli, and takes on 1 with probability p, and 0 with probability $1-p$. Then the mgf of X is

$$\begin{aligned}
E[e^{tX}] &= \sum_{x=0}^{1} e^{tx} P(X = x) \\
&= e^{t0}(1-p) + e^t p \\
&= 1 - p + pe^t.
\end{aligned}$$

Example 16. Let X be the uniform random variable over $[a, b]$. Then the mgf of X is

$$\begin{aligned}
E[e^{tX}] &= \int_a^b e^{tx} \frac{1}{b-a}\, dx \\
&= \left. \frac{e^{tx}}{t(b-a)} \right|_{x=a}^{x=b} \\
&= \frac{e^{tb} - e^{ta}}{t(b-a)}.
\end{aligned}$$

The moment generating function, from its name, should involve the moments. Indeed, we have that

$$E[X^n] = \left. \frac{d^n}{dt^n} \right|_{t=0} [M_X(t)].$$

That is, we differentiate the mgf n times and evaluate $t = 0$ and that will give us the nth moment.

Example 17. We shall use the mgf of a Bernoulli random variable to find its mean and variance. Using that the mgf is $M_X(t) = 1 - p + pe^t$, we have

$$E[X] = \left. \frac{d}{dt} \right|_{t=0} [1 - p + pe^t] = (pe^t)|_{t=0} = p.$$

To find its variance, we need to compute

$$\begin{aligned}
E[X^2] &= \left. \frac{d^2}{dt^2} \right|_{t=0} [1 - p + pe^t] \\
&= \left. \frac{d}{dt} \right|_{t=0} [pe^t] \\
&= p.
\end{aligned}$$

Then $\text{Var}(X) = E[X^2] - E[X]^2 = p - p^2 = p(1 = p)$.

Sometimes though, differentiation can be annoying. Another way to get the moments is to write out the mgf as a power series, since we have that

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

Example 18. In the case of the uniform random variable, differentiating the mgf is annoying. So we write the mgf as an infinite series

$$\begin{aligned}
\frac{e^{tb}-e^{ta}}{t(b-a)} &= \frac{1}{t(b-a)}\left[1+tb+\frac{(tb)^2}{2!}+\dots\right.\\
&\qquad\left.-\left(1+ta+\frac{(ta)^2}{2!}+\dots\right)\right]\\
&= \frac{1}{t(b-a)}\left[(tb-ta)+\frac{(tb)^2-(ta)^2}{2!}+\dots\right]\\
&= \left[\frac{b-a}{b-a}+\frac{(b^2-a^2)t}{(b-a)2\times 1!}+\frac{(b^3-a^3)t^2}{(b-a)3\times 2!}+\dots\right]\\
&= \sum_{n=0}^{\infty}\left(\frac{b^n-a^n}{(n+1)(b-a)}\right)\frac{t^n}{n!}.
\end{aligned}$$

Therefore, for the uniform random variable, $E[X^n]=\frac{b^n-a^n}{(n+1)(b-a)}$.

Let X and Y be independent random variables. Then the distribution of the sum of X and Y can be easily computed using mgfs, as we have

$$\boxed{M_{X+Y}(t)=M_X(t)M_Y(t).}$$

Once we find the mgf of a sum of independent random variables, we find that it matches a family of mgfs, thus $X+Y$ must have such a distribution.

Example 19. Let X be binomial with parameters n, p. If we were to compute the mgf of the binomial directly, it would be difficult. Instead, we can write $X=X_1+X_2+\dots+X_n$, where each X_i is Bernoulli. Then we have that

$$\begin{aligned}
M_X(t) &= M_{X_1+X_2+\dots+X_n}(t)\\
&= M_{X_1}(t)\times M_{X_2}(t)\times\dots\times M_{X_n}(t)\\
&= (1-p+pe^t)^n.
\end{aligned}$$

Example 20. Let X and Y be normal with mean μ and variance σ^2. The mgf of the normal is $e^{\mu t+\frac{1}{2}\sigma^2t^2}$. Then we have that

$$\begin{aligned}
M_{X+Y}(t) &= M_X(t)M_Y(t)\\
&= \left(e^{\mu t+\frac{1}{2}\sigma^2t^2}\right)\left(e^{\mu t+\frac{1}{2}\sigma^2t^2}\right)\\
&= e^{2\mu t+\frac{1}{2}2\sigma^2t^2}.
\end{aligned}$$

Therefore, $X+Y$ is a normal random variable with mean 2μ and variance $2\sigma^2$.

(*)We note that the moments of a random variable do not need to exist, and indeed, the Cauchy distribution is an example where none of the moments exist. In such a case, the moment generating function does not exist. However, if we start to include complex numbers, a variation of the mgf called the **characteristic function**, defined to be $E[e^{itX}]$, is guaranteed to always exist.

1.4 Limit Theorems

1.4.1 Markov's Inequality

Markov's inequality states that if a nonnegative random variable has a small mean, then the probability for it to take a large value is small.

Theorem 1 (Markov's inequality)**.** Let X be a nonnegative random variable. Then for any $a > 0$, we have

$$\boxed{P(X > a) < \frac{E[X]}{a}.}$$

1.4.2 Chebyshev's Inequality

Chebyshev's inequality states that if a random variable has a small variance, then the probability it takes a value far from the mean is small.

Theorem 2 (Chebyshev's inequality)**.** Let X be a random variable with mean μ and variance σ^2. Then

$$\boxed{P(|X - \mu| > c) \leq \frac{\sigma^2}{c^2}.}$$

1.4.3 The Weak Law of Large Numbers

Let $X_1, \ldots, X_n$ be independent and identically distributed. The **sample mean** is defined to be

$$\overline{X}_n = \frac{X_1 + \cdots + X_n}{n}.$$

The **weak law of large numbers** says the sample mean is close to the true mean with high probability.

Theorem 3 (Weak Law of Large Numbers)**.** Let $X_1, X_2, \ldots$ be independent and identically distributed random variables with mean μ. For every $\epsilon > 0$, we have

$$\boxed{\lim_{n\to\infty} P\left(\left|\frac{X_1 + X_2 + \cdots + X_n}{n} - \mu\right| < \epsilon\right) = 1.}$$

1.4.4 The Central Limit Theorem

Perhaps the most important theorem in probability is the central limit theorem. It states that the sum of independent random variables approaches the normal distribution, even if the original random variables are not normally distributed. In real life problems, it explains why many events follow a bell shaped curve.

Theorem 4 (Central Limit Theorem)**.** Let $X_1, X_2, \ldots$ be independent and identically distributed with common mean μ and variance σ^2. Define

$$Z_n = \frac{X_1 + \cdots + X_n - n\mu}{\sigma\sqrt{n}}.$$

1. The **sum form** of the central limit theorem states that the CDF of Z_n converges to the CDF of the standard normal

   $$\Phi(z) = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2}\, dz.$$

   That is,

   $$\boxed{\lim_{n\to\infty} P\left(\frac{X_1 + \ldots X_n - n\mu}{\sigma\sqrt{n}} \leq z\right) = \Phi(z).}$$

2. Equivalently, the **mean form** of the central limit theorem states

   $$\boxed{\lim_{n\to\infty} P\left(\frac{\frac{X_1 + \ldots X_n}{n} - \mu}{\sigma/\sqrt{n}} \leq z\right) = \Phi(z).}$$

   That is, the cdf of the normalized sample mean approaches the cdf of the standard normal.

Now we are able to use the CLT in approximating probabilities. Looking carefully at the CLT, we no longer need to deal with working with pmfs or pdfs to compute probabilities, and instead only need to work with a Z-table, as well as some knowledge of working with means and variances.

Using the Central Limit Theorem

- If a problem asks for the probability related to a sum, let $S_n = X_1 + \cdots + X_n$, where the X_i are independent and identically distributed random variables with mean μ and variance σ^2. If n is large, the probability $P(S_n < c)$ can be approximated by treating S_n as if it were normal, according to the following procedure.
    1. Calculate the mean $n\mu$ and the variance $n\sigma^2$ of S_n.
    2. Calculate the normalized value $z = \frac{c-n\mu}{n\sigma^2}$.
    3. Use the approximation $P(S_n < c) \approx \Phi(z)$, where $\Phi(z)$ is available from standard normal CDF tables.
- In the case that a problem asks for the probability related to a mean, let $\overline{X}_n = \frac{X_1+\cdots+X_n}{n}$. Then to find $P(\overline{X}_n < c)$, we do the following procedure.
    1. Calculate the mean μ and the variance $\sigma^2/\sqrt{n}$ of $\overline{X}_n$.
    2. Calculate the normalized value $z = \frac{c-\mu}{\sigma^2/\sqrt{n}}$.
    3. Use the approximation $P(\overline{X}_n < c) \approx \Phi(z)$, where $\Phi(z)$ is available from standard normal CDF tables.

Example 21. Let $X_1, X_2, \ldots$ be independent and identically distributed uniform random variables over $[5, 50]$. Find the probability that $X_1 + \cdots + X_{100} > 3000$. *Solution:* Let $S_{100} = X_1 + \cdots + X_{100}$. We want to find $P(S_{100} > 3000)$. FIrst, we need to compute $E[X_1]$ and $\text{Var}(X_1)$. We get

$$E[X_1] = \frac{5+50}{2} = 27.5, \qquad \text{Var}(X_1) = \frac{(50-5)^2}{12} = 168.75.$$

Then we have that

$$\begin{aligned} P(S_{100} > 3000) &= P\left(\frac{S_{100} - 100 \times 27.5}{\sqrt{100 \times 168.75}} > \frac{3000 - 100 \times 27.5}{\sqrt{100 \times 168.75}}\right) \\ &\approx P(Z > 1.92) \\ &= 1 - \Phi(1.92) \\ &= 1 - 0.9726 = 0.0274. \end{aligned}$$

1.4.5 The Strong Law of Large Numbers

The weak law of large numbers states that there is a high probability that the sample mean approaches the actual mean, but it does not say whether or not the sample mean itself converges to the actual mean. The **strong law of large numbers** says that the sample mean indeed converges to the true mean except with a technicality.

Thus what we want is this. Let $X_1, X_2, \ldots$ be independent and identically distributed random variables with mean μ. Then the sequence of sample means $\overline{X}_n$ converges to μ, meaning that

$$\lim_{n\to\infty} \frac{X_1 + X_2 + \cdots + X_n}{n} = \mu.$$

Unfortunately, this is wrong.

In the world of random variables, we do not care about sets that have probability zero of happening. The result of this is that it is possible that sample mean does not converge to the true mean on a set of probability 0. Thus, the sample mean converges to the true mean, except possibly on a set of probability 0. But we should emphasize that the strong law states the sample mean converges to the true mean, and with probability 1 is a technicality.

Theorem 5 (Strong Law of Large Numbers)**.** Let $X_1, X_2, \ldots$ be independent and identically distributed random variables with mean μ. Then the sequence of sample means $\overline{X}_n$ converges to μ with probability 1, meaning that

$$\boxed{P\left(\lim_{n\to\infty} \frac{X_1 + X_2 + \cdots + X_n}{n} = \mu\right) = 1.}$$

2 Summary

Here are the essential topics we should know. Do at least one or two problems for each.

1. What is the probability density function, and what is its relationship with the cumulative distribution function?
2. Know how to problems involving the uniform distribution.
3. Know how to use a Z-table. When a problem involves a normal distribution, how do we convert it to a normal distribution with mean 0 and standard deviation 1?
4. How do we approximate a binomial with the normal? How do we adjust for the fact that a binomial is discrete, and the normal is continuous?
5. Know how to do problems involving exponential distributions. When there is a conditional probability, it is probably convenient to use the memoryless property.
6. Know the relationship of two or more random variables. How do we find the probability of events like $P(X < Y)$?
7. How do we find the marginal distributions. Make sure the marginal density functions are complete.
8. How do we know that two random variables are independent?
9. Know how to use linearity of expectation in problems, splitting up random variables into sums.
10. Know the formula for both covariance and the correlation coefficient.

3 Review Problems

3.1 Continuous Random Variables

1. Let X be a continuous random variable with probability density function

$$f(x) = \begin{cases} c(1-x^2) & -1 < x < 1 \\ 0 & \text{otherwise,} \end{cases}$$

for some value of c. What is the value of c? What is the cumulative distribution function of X?

2. Let X be a random variable with probability density function

$$f(x) = \begin{cases} \frac{4}{\pi}\sqrt{1-x^2} & 0 < x < 1 \\ 0 & \text{otherwise.} \end{cases}$$

Find $E[X]$.

3. Let X have the pdf

$$f(x) = \frac{\lambda}{2} e^{-\lambda|x|},$$

where $\lambda > 0$. Show that $P(-\infty < X < \infty) = 1$. Find the mean and variance of X.

4. Trains headed for destination A leave at 15-minute intervals starting at 7am, whereas trains headed for destination B leave at 15 minute intervals starting at 7:05am.
   (a) If a certain passenger arrives at the station at a time uniformly distributed between 7 and 8am, and then gets on the first train that arrives, what proportion of time does he or she go to destination A?
   (b) What if the passenger arrives at a time uniformly distributed between 7:10 and 8:10am?

5. A point is chosen at random on a line segment of length L. Interpret this statement, and find the probability that the ratio of the shorter to the longer segment is less than $\frac{1}{4}$.

6. You arrive at a bus stop at 10am, knowing that the bus will arrive at some time uniformly distributed between 10 and 10:30.
   (a) What is the probability that you will have to wait longer than 10 minutes?
   (b) If, at 10:15, the bus has not yet arrived, what is the probability that you will have to wait at least an additional 10 minutes?

7. If X is a normal random variable with parameters $\mu = 10$ and $\sigma^2 = 36$, compute
   (a) $P(X > 5)$,
   (b) $P(4 < X < 16)$,
   (c) $P(X < 8)$,
   (d) $P(X < 20)$,
   (e) $P(X > 16)$.

8. Suppose that X is a normal random variable with mean 5. If $P(X > 9) = 0.2$, approximately what is $\mathrm{Var}(X)$?

9. Let X be a normal random variable with mean 12 and variance 4. Find the value of c such that $P(X > c) = 0.10$.

10. One thousand independent rolls of a fair die will be made. Compute an approximation to the probability that the number 6 will appear between 150 and 200 times inclusively. If the number 6 appears exactly 200 times, find the probability that the number 5 will appear less than 150 times.

11. Show that if Z is a standard normal random variable, then for $x > 0$,
    (a) $P(Z > x) = P(Z < -x)$,
    (b) $P(|Z| > x) = 2P(Z > x)$,
    (c) $P(|Z| < x) = 2P(Z < x) - 1$.

12. The time in hours to repair a machine is an exponentially distributed random variable with $\lambda = \frac{1}{2}$.
    (a) What is the probability that a repair time exceeds 2 hours?
    (b) What is the conditional probability that a repair takes at least 10 hours, given that its duration exceeds 9 hours?

13. The number of years a radio functions is exponentially distributed with parameter $\lambda = \frac{1}{8}$. If Jones buys a used radio, what is the probability that it will be working after an additional 8 years?

14. At a certain bank, the amount of time that a customer spends being served by a teller is an exponential random variable with mean 5 minutes. If there is a customer in service when you enter the bank, what is the probability that he or she will still be with the teller after an additional 4 minutes?

15. If X is an exponential random variable with mean $1/\lambda$, and $c > 0$, show that cX is exponential with mean c/λ.

16. Let X be a continuous random variable, with cumulative distribution function $F_X(x)$ and probability density function $f_X(x)$. Let $Y = 5X - 2$. Find an expression for the probability density function $f_Y(y)$ of Y.

17. Let X be a normal random variable with mean 0 and variance 1. Show that $Y = X^2$ is a gamma random variable. Note that $y = x^2$ is not injective. In particular, X^2 is a special case of the gamma distribution called the **chi-squared** distribution, of interest in statistics.

3.2 Jointly Distributed Random Variables

18. Flip a fair coin twice. Let X be 1 if the first coin is heads, and 0 if the first coin is tails. Let Y be the total number of heads. Find the joint mass function of X and Y. Find the marginal mass functions of X and Y.

19. Let X and Y be independent uniform distributions on $(0, 1)$. What is $P(X < Y)$?

20. Let X and Y have the joint density function

$$f(x, y) = \begin{cases} \frac{1}{2}e^{-(x+\frac{1}{2}y)} & 0 \le x < \infty, \quad 0 \le y < \infty \\ 0 & \text{otherwise.} \end{cases}$$

    Compute the marginal density of X and the marginal density of Y(don't forget the bounds). Are X and Y independent? Compute $P(X < Y)$.

21. The joint probability density function of X and Y is given by

$$f(x, y) = c(y^2 - x^2)e^{-y} \quad -y \le x \le y, 0 < y < \infty.$$

    (a) Find c.
    (b) Find the marginal densities of X and Y.
    (c) Find $E[X]$.

22. A man and a woman agree to meet at a certain location around 12:30pm. If the man arrives at a time uniformly distributed between 12:15 and 12:45, and if the woman independently arrives at a time uniformly distributed between 12:00 and 1:00pm, find the probability that the first to arrive waits no longer than 5 minutes. What is the probability that the man arrives first?

23. An ambulance travels back and forth at a constant speed along a road of length L. At a certain moment of time, an accident occurs at a point uniformly distributed on the road. Assuming the ambulance's location is also uniformly distributed, and assuming the independence of the variables, compute the distribution of the distance of the ambulances from the accident.
Hint: Let X and Y be uniform random variables. Compute $P(|X - Y| < a)$. Drawing a picture may help.

24. Let X and Y have the joint density function

$$f(x, y) = \begin{cases} 6x & 0 < x < y < 1 \\ 0 & \text{otherwise.} \end{cases}$$

Compute the marginal density of X and the marginal density of Y(don't forget the bounds). Compute $P(X + Y < 1)$.
Hint: Be careful with the bounds of integration. It may be helpful to draw a picture of the domain of integration.

25. Let

$$f(x, y) = \begin{cases} 24xy & 0 \leq x \leq 1, 0 \leq y \leq 1, 0 \leq x + y \leq 1, \\ 0 & \text{otherwise.} \end{cases}$$

   (a) Show that $f(x, y)$ is a joint probability density function.
   (b) Find $E[X]$ and $E[Y]$.

26. The joint density function of X and Y is

$$f(x, y) = \begin{cases} xy & 0 < x < 1, 0 < y < 2, \\ 0 & \text{otherwise.} \end{cases}$$

   (a) Are X and Y independent?
   (b) Find the density function of X.
   (c) Find the density function of Y.
   (d) Find the joint distribution function.
   (e) Find $E[Y]$.
   (f) Find $P(X + Y < 1)$.

27. Let X be exponential with mean $1/\lambda$. Find $P([X] = n, X - [X] \leq x)$, where $[x]$ is defined to be the largest integer less than or equal to x. Conclude that $[X]$ and $X - [X]$ are independent.

28. Let X, Y, and Z be independent random variables having identical density functions $f(x) = e^{-x}, 0 < x < \infty$. Derive the joint distribution of $U = X + Y, V = X + Z, W = Y + Z$.

3.3 Properties of Expectation

29. A player throws a fair die and simultaneously flips a fair coin. If the coin lands heads, then she wins twice the value that appears on the die. If the coin lands tails, then she wins half the value that appears on the die. Determine her expected winnings.

30. A fair die is rolled 10 times. Calculated the expected sum of the 10 rolls.

31. Consider n independent flips of a coin with probability p of landing on heads. Say that a changeover occurs whenever an outcome differs from the one preceding it. For instance, if $n = 5$ and the outcome is $HHTHT$, then there are 3 changeovers. Find the expected number of changeovers.

32. If X and Y are independent and identically distributed with mean μ and variance σ^2, find

$$E[(X-Y)^2].$$

33. If $E[X]=1$ and $\mathrm{Var}(X)=5$, find

   (a) $E[(2+X)^2]$,

   (b) $\mathrm{Var}(4+3X)$.

34. Let X and Y have a joint density function given by

$$f(x,y)=\begin{cases} 2e^{-2x}/x & -<x<\infty, 0<y<x, \\ 0 & \text{otherwise.} \end{cases}$$

   Compute $\mathrm{Cov}(X,Y)$.

4 More Difficult Problems

The following problems are more difficult. It is recommended to do these problems after doing the problems in the previous section.

1. Let X be a non-negative random variable, and suppose that $E[X^n] < \infty$ for all n. Show that

$$E[X^n] = \int_0^\infty nx^{n-1} P(X > x)\, dx.$$

2. Calamity Jane goes to the bank to make a withdrawal, and is equally likely to find 0 or 1 customers ahead of her. The service time of the customer ahead, if present, is exponentially distributed with mean $1/\lambda$. What is the cdf of Jane's waiting time?

3. Alvin throws darts at a circular target of radius r and is equally likely to hit any point in the target. Let X be the distance of Alvin's hit from the center.
   (a) Find the PDF, the mean, and the variance of X.
   (b) The target has an inner circle of radius t. If $X \leq t$, Alvin gets a score of $S = 1/X$. Otherwise his score is $S = 0$. Find the CDF of S. Is S a continuous random variable?

4. Let X, Y, Z be iid uniform from 0 to 1. Find $P(XY < Z^2)$.

5. A point is chosen at random (according to a uniform PDF) within a semicircle of the form $\{(x, y) \mid x^2 + y^2 \leq r^2, y \geq 0\}$, for some given $r > 0$.
   (a) Find the joint PDF of the coordinates X and Y of the chosen point.
   (b) Find the marginal PDF of Y and use it to find $E[Y]$.
   (c) Check your answer in (b) by computing $E[Y]$ directly without using the marginal PDF of Y.

6. A stick of length 1 is broken into two at random.
   (a) What is the average length of the smaller piece?
   (b) What is the average ratio of the smaller piece to the larger piece?

7. We start with a stick of length ℓ. We break it at a point which is chosen according to a uniform distribution and keep the piece, of length Y, that contains the left end of the stick. We then repeat the same process on the piece that we were left with, and let X be the length of the remaining left piece after breaking for the second time.
   (a) Find the joint pdf of Y and X.
   (b) Find the marginal pdf of X.
   (c) Use the pdf of X to evaluate $E[X]$.
   (d) Evaluate $E[X]$ by exploiting the relation $X = Y \times \left(\frac{X}{Y}\right)$.

8. The coordinates X and Y of a point are independent normal random variables with mean 0 and common variance σ^2. Given that the point is at a distance of at least c from the origin, find the conditional joint PDF of X and Y.

9. Let X and Y have the joint density function

$$f(x, y) = \begin{cases} 12xy(1-y) & 0 < x < 1, 0 < y < 1, \\ 0 & \text{otherwise.} \end{cases}$$

   Let $Z = XY^2$. Find the joint density of Y and Z. Use the joint density to find the marginal density of Z.

10. Let $X_1, X_2, \ldots, X_n$ be independent and identically distributed random variables with common pdf $f(x)$. Let $Y_1, Y_2, \ldots, Y_n$ be the order statistics. Provide an argument that the joint pdf of the order statistics is

$$f_{Y_1,\ldots,Y_n}(y_1, \ldots, y_n) = \begin{cases} n! \prod_{k=1}^{n} f(y_k) & y_1 < y_2 < \cdots < y_n \\ 0 & \text{otherwise.} \end{cases}.$$

11. By integrating the joint pdf of the order statistics, find the marginal density of the first order statistic Y_1, and the nth order statistic Y_n.

12. For $1 \leq k \leq n$, find the marginal pdf of the kth order statistic.

13. Let X, Y have the joint pdf

$$f(x, y) = \begin{cases} 2x^2 y + \sqrt{y} & 0 < x < 1, 0 < y < 1 \\ 0 & \text{otherwise.} \end{cases}$$

Let $U = \min\{X, Y\}$ and $V = \max\{X, Y\}$. Find the joint pdf of U and V.

Note: X and Y are not independent and identically distributed, so order statistics will not apply here.

14. Let X and Y be two independent random variables. X is exponentially distributed with mean 1, and Y is uniformly distributed on $[0, 1]$. Calculate $E[\max\{X, Y\}]$.

15. A point (X, Y) is uniformly distributed on the unit square $[0, 1]^2$. Let Θ be the angle between the x-axis and the line segment that connects $(0, 0)$ to the point (X, Y). Find the expected value $E[\Theta]$.

16. Let X and Y be jointly continuous with joint density function $f_{X,Y}(x, y) = \frac{1}{x}, 0 \leq y \leq x \leq 1$. Compute the probability $P(X^2 + Y^2 \leq 1 \mid X = x)$.

17. Let X and Y be normal distributions(but not necessarily independent) with common mean 0 and common variance 1. For $-1 \leq \rho \leq 1$, we have that X and Y follow a **bivariate normal distribution**. The joint density function of X and Y is

$$f(x, y) = \frac{1}{2\pi\sqrt{1-\rho^2}} e^{-\frac{1}{2}\frac{x^2+y^2-2\rho xy}{(1-\rho^2)}} \qquad -\infty < x < \infty, -\infty < y < \infty.$$

Define

$$X^+ = \begin{cases} X & \text{if } X > 0 \\ 0 & \text{if } X \leq 0. \end{cases}$$

Find $E[X^+]$. Find $E[X^+ \mid Y = y]$.

A List of Moment Generating Functions

Discrete Random Variables

Distribution	Mean	Variance	Moment Generating Function $M(t)$
Bernoulli	p	$p(1-p)$	$pe^t + 1 - p$
Binomial	np	$np(1-p)$	$(pe^t + 1 - p)^n$
Poisson	λ	λ	$e^{\lambda(e^t - 1)}$
Geometric	$\frac{1}{p}$	$\frac{1-p}{p^2}$	$\frac{pe^t}{1-(1-p)e^t}$
Negative Binomial	$\frac{r}{p}$	$\frac{r(1-p)}{p^2}$	$\left[\frac{pe^t}{1-(1-p)e^t}\right]^r$

Continuous Random Variables

Distribution	Mean	Variance	Moment Generating Function $M(t)$
Uniform	$\frac{a+b}{2}$	$\frac{(b-a)^2}{12}$	$\frac{e^{tb} - e^{tb}}{t(b-a)}$
Normal	μ	σ^2	$e^{\mu t + \frac{1}{2}\sigma^2 t^2}$
Exponential	$\frac{1}{\lambda}$	$\frac{1}{\lambda^2}$	$\frac{\lambda}{\lambda - t}$
Gamma	$\frac{\alpha}{\lambda}$	$\frac{\alpha}{\lambda^2}$	$\left[\frac{\lambda}{\lambda - t}\right]^\alpha$
Cauchy	DNE	DNE	DNE

B Sum of Independent Random Variables

Let X and Y be independent random variables. Let $Z = X + Y$. Then to get the cumulative distribution function of Z, we have

$$\begin{aligned} P(X+Y<z) &= \iint_{x+y<z} f_X(x) f_Y(y) dx dy \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{z-y} f_X(x) f_Y(y) dx dy \\ &= \int_{-\infty}^{\infty} F_X(z-y) f_Y(y) dy. \end{aligned}$$

To find the probability density function, we differentiate with respect to z and get

$$\begin{aligned} f_{X+Y}(z) &= \frac{d}{dz} \int_{-\infty}^{\infty} F_X(z-y) f_Y(y) dy \\ &= \int_{-\infty}^{\infty} \frac{d}{dz} [F_X(z-y) f_Y(y)] dy \\ &= \int_{-\infty}^{\infty} f_X(z-y) f_Y(y) dy. \end{aligned}$$

We shall now apply this formula to find the pdf of the sum of X and Y, where both are uniform random variables from 0 to 1.
Let us define the **indicator** function to be

$$\mathbf{1}_E(x) = \begin{cases} 1 & x \in E \\ 0 & x \notin E. \end{cases}$$

Then the probability density function of X can be written with the indicator function to be

$$f_X(x) = \mathbf{1}_{[0,1]}(x) \quad -\infty < x \infty.$$

Then we have

$$f_{X+Y}(z) = \int_{-\infty}^{\infty} \mathbf{1}_{[0,1]}(z-y) \mathbf{1}_{[0,1]}(y) dy.$$

Then as a function of z, we see that

$$\mathbf{1}_{[0,1]}(z-y)\mathbf{1}_{[0,1]}(y) = \begin{cases} 1 & 0 \leq (z-y) \leq 1, 0 \leq y \leq 1 \\ 0 & \text{otherwise.} \end{cases}$$

Then $0 \leq (z-y) \leq 1$ is equivalent to $y \leq z \leq 1 + y$. We want to combine the two inequalities $y \leq z \leq 1 + y$ and $0 \leq y \leq 1$. But combining the two inequalities depends on the value of z, so we analyze the different cases.

1. Suppose $z < 0$. We have that $y \leq z < 0$, so $y < 0$, and we are given $0 \leq y \leq 1$. Then no values of y can satisfy both $y < 0$ and $y \geq 0$. Then the indicator function is 0.

2. Suppose $0 \leq z < 1$. We have that $y < z < 1$ and $0 < y < 1$. Putting the two inequalities together, we have that $0 < y < z < 1$. So the indicator function is 1 when $0 < z < 1$ and $0 < y < z$.

3. Suppose $1 < z < 2$. We have that $1 < z < y + 1$ and $0 < y < 1$. So we get that the indicator function is 1 when $z - 1 < y < 1$.

4. Suppose $z > 2$. Then $2 < z < 1 + y$, so $1 < y$. But at the same time, $0 < y < 1$, so no values of y can satisfy both inequalities at the same time. Then the indicator function is 0.

Putting it together, we get

$$\mathbf{1}_{[0,1]}(z-y)\mathbf{1}_{[0,1]}(y) = \begin{cases} 1 & \{0 < z < 1\} \text{ and } \{0 < y < z\} \\ 1 & \{1 < z < 2\} \text{ and } \{z-1 < y < 1\} \\ 0 & \{z < 0\} \text{ or } \{z > 2\}. \end{cases}$$

Then we can evaluate the integral,

1. When $0 < z < 1$, we force $0 < y < z$ in order for the indicator function to be 1, and so

$$\begin{aligned} f_{X+Y}(z) &= \int_{-\infty}^{\infty} \mathbf{1}_{[0,1]}(z-y)\mathbf{1}_{[0,1]}(y)dy \\ &= \int_{0}^{z} 1dy \\ &= z. \end{aligned}$$

2. When $1 < z < 2$, we force $z - 1 < y < 1$ in order for the indicator function to be 1, and so

$$\begin{aligned} f_{X+Y}(z) &= \int_{-\infty}^{\infty} \mathbf{1}_{[0,1]}(z-y)\mathbf{1}_{[0,1]}(y)dy \\ &= \int_{z-1}^{1} 1dy \\ &= 2 - z. \end{aligned}$$

3. When $z < 0$ or $z > 2$, we have that the indicator function is 0 for all y, so $f_{X+Y}(z) = 0$.

Putting it all together, we get that the probability density function of the sum of two uniform random variables from 0 to 1 is

$$f_{X+Y}(z) = \begin{cases} z & 0 < z < 1 \\ 2 - z & 1 < z < 2 \\ 0 & \text{otherwise.} \end{cases}$$

C Order Statistics

Let $X_1, X_2, \dots, X_n$ be independent and identically distributed continuous random variables. Let $F(x)$ be the cdf of X_k, and let $f(x)$ be the pdf of X_k. We define the *order statistics* $Y_1, \dots, Y_n$ to be

$$\begin{aligned}
Y_1 &= \text{smallest of } X_1, \dots, X_n, \\
Y_2 &= \text{second smallest of } X_1, \dots, X_n, \\
&\vdots \\
Y_j &= j\text{th smallest of } X_1, \dots, X_n, \\
&\vdots \\
Y_n &= \text{largest of } X_1, \dots, X_n.
\end{aligned}$$

Now let us find the joint cdf of the order statistics. Let us work concretely, and let $n = 3$. Consider

$$P(Y_1 < y_1, Y_2 < y_2, Y_3 < y_3).$$

In order for this to happen, we have one of the following situations to happen:

$$\begin{aligned}
&P(X_1 < y_1, X_2 < y_2, X_3 < y_3), && P(X_1 < y_1, X_3 < y_2, X_2 < y_3), \\
&P(X_2 < y_1, X_1 < y_2, X_3 < y_3), && P(X_2 < y_1, X_3 < y_2, X_1 < y_3), \\
&P(X_3 < y_1, X_1 < y_2, X_2 < y_3), && P(X_3 < y_1, X_2 < y_2, X_1 < y_3).
\end{aligned}$$

The situations are mutually exclusive, and each corresponds to one of the 3! permutations of X_1, X_2, X_3 that can happen. Since X_1, X_2, X_3 are independent, we can split up the probabilities into the product of the cdfs of X_i, and since they are identically distributed, they have the same cdfs. This means $P(X_1 < y_1, X_2 < y_2, X_3 < y_3) = F(y_1)F(y_2)F(y_3)$. This is true for each of the situations above. So in conclusion, we get

$$\begin{aligned}
P(Y_1 < y_1, Y_2 < y_2, Y_3 < y_3) &= P(X_1 < y_1, X_2 < y_2, X_3 < y_3) + P(X_1 < y_1, X_3 < y_2, X_2 < y_3) \\
&\quad + P(X_2 < y_1, X_1 < y_2, X_3 < y_3) + P(X_2 < y_1, X_3 < y_2, X_1 < y_3) \\
&\quad + P(X_3 < y_1, X_1 < y_2, X_2 < y_3) + P(X_3 < y_1, X_2 < y_2, X_1 < y_3) \\
&= F(y_1)F(y_2)F(y_3) + F(y_1)F(y_2)F(y_3) \\
&\quad + F(y_1)F(y_2)F(y_3) + F(y_1)F(y_2)F(y_3) \\
&\quad + F(y_1)F(y_2)F(y_3) + F(y_1)F(y_2)F(y_3) \\
&= 3!F(y_1)F(y_2)F(y_3).
\end{aligned}$$

There was nothing special about $n = 3$, so in general, we have $n!$ permutations that we need to work with. So we conclude that the *joint cdf of the order statistics* is

$$F(y_1, y_2, \dots, y_n) = n! \prod_{i=1}^{n} F(y_i).$$

Taking partial derivatives, we get that the *joint pdf of the order statistics* is

$$f(y_1, y_2, \dots, y_n) = \begin{cases} n! \prod_{i=1}^{n} f(y_i) & y_1 < y_2 < \dots < y_n, \\ 0 & \text{otherwise.} \end{cases}$$

Now the next step is to find the marginal densities of each of the order statistics. Again for concreteness, let us work in the case $n = 3$. Then first, let us find the marginal density of Y_1. In this case, the best order of integration is the highest to lowest, which here is $dy_3 dy_2$. We have that

$$\begin{aligned}
f_{Y_1}(y_1) &= \int_{y_1}^{\infty}\int_{y_2}^{\infty} 3! f(y_1)f(y_2)f(y_3)dy_3dy_2 \\
&= 3! f(y_1)\int_{y_1}^{\infty} f(y_2)\int_{y_2}^{\infty} f(y_2)f(y_3)dy_3dy_2 \\
&= 3! f(y_1)\int_{y_1}^{\infty} f(y_2)[F(y_3)]_{y_2}^{\infty}dy_2 \\
&= 3! f(y_1)\int_{y_1}^{\infty} f(y_2)[1-F(y_2)]dy_2 \\
&= \frac{3!}{2} f(y_1)[-(1-F(y_2))^2]_{y_1}^{\infty} \\
&= \frac{3!}{2} f(y_1)[1-F(y_1)]^2.
\end{aligned}$$

Now we find the marginal density of Y_3.

$$\begin{aligned}
f_{Y_3}(y_3) &= \int_{-\infty}^{y_3}\int_{-\infty}^{y_2} 3! f(y_1)f(y_2)f(y_3)dy_1dy_2 \\
&= 3! f(y_3)\int_{-\infty}^{y_3}\int_{-\infty}^{y_2} f(y_1)f(y_2)dy_1dy_2 \\
&= 3! f(y_3)\int_{-\infty}^{y_3} f(y_2)[F(y_1)]_{-\infty}^{y_2}dy_2 \\
&= 3! f(y_3)\int_{-\infty}^{y_3} f(y_2)F(y_2)dy_2 \\
&= \frac{3!}{2} f(y_3)[F(y_2)^2]_{-\infty}^{y_3} \\
&= \frac{3!}{2} f(y_3)[F(y_3)]^2.
\end{aligned}$$

Then for the marginal density of Y_2, we get

$$\begin{aligned}
f_{Y_2}(y_2) &= \int_{-\infty}^{y_2}\int_{y_2}^{\infty} 3! f(y_1)f(y_2)f(y_3)dy_1dy_3 \\
&= 3! f(y_2)\int_{-\infty}^{y_2} f(y_1)dy_1\int_{y_2}^{\infty} f(y_3)dy_3 \\
&= 3! F(y_2)f(y_2)[1-F(y_2)].
\end{aligned}$$

Exercise: Compute by the means of integration the marginal densities of the order statistics when $n = 4$. Then for general n, all we do is integrate more, but it follows the same pattern. So the marginal density of the kth order statistic Y_k is

$$f_{Y_k}(y_k) = \binom{n}{k-1, 1, n-k} F(y_k)^{k-1} f(y_k)[1-F(y_k)]^{n-k}.$$

D Deriving the Formula for Conditional Distributions

Let X and Y be two discrete random variables. Remember that the conditional probability of E given F is $P(E|F) = P(E \cap F)/P(F)$. Using this formula, the **conditional mass function** of X given $Y = y$ should be

$$P(X = x|Y = y) = \frac{P(\{X = x\} \cap \{Y = y\})}{P(Y = y)}.$$

Now suppose that X and Y are two continuous random variables. Looking at the discrete case, the formula for the conditional probability density function of X given $Y = y$ should be

$$f_{X|Y}(x|y) = \frac{f_{X,Y}(x, y)}{f_Y(y)}.$$

But this doesn't follow directly from the definition of conditional probability. If we try to use the formula for conditional probability given above, the conditional cumulative distribution function of X given $Y = y$ should be

$$P(X < x|Y = y) = \frac{P(\{X < x\} \cap \{Y = y\})}{P(Y = y)}.$$

However in this case, since Y is continuous, we have that both the numerator and the denominator are equal to zero. Instead, let us have an epsilon of room, so let $\epsilon > 0$, and we define

$$P(X < x|Y = y) = \lim_{\epsilon \to 0} \frac{P(\{X < x\} \cap \{y < Y < y + \epsilon\})}{P(y < Y < y + \epsilon)}.$$

We do some algebraic manipulations to get

$$\begin{aligned}
\lim_{\epsilon \to 0} \frac{P(\{X < x\} \cap \{y < Y < y + \epsilon\})}{P(y < Y < y + \epsilon)} &= \lim_{\epsilon \to 0} \frac{P(X < x, Y < y + \epsilon) - P(X < x, Y < y)}{P(Y < y + \epsilon) - P(Y < y)} \\
&= \lim_{\epsilon \to 0} \frac{\left(\dfrac{P(X < x, Y < y + \epsilon) - P(X < x, Y < y)}{\epsilon}\right)}{\left(\dfrac{P(Y < y + \epsilon) - P(Y < y)}{\epsilon}\right)} \\
&= \frac{\lim_{\epsilon \to 0} \left(\dfrac{F_{X,Y}(x, y + \epsilon) - F_{X,Y}(x, y)}{\epsilon}\right)}{\lim_{\epsilon \to 0} \left(\dfrac{F_Y(y + \epsilon) - F_Y(y)}{\epsilon}\right)} \\
&= \frac{\frac{\partial}{\partial y}[F_{X,Y}(x, y)]}{f_Y(y)}.
\end{aligned}$$

So the conclusion of the calculation is that the conditional cumulative distribution function is

$$P(X < x|Y = y) = \frac{\frac{\partial}{\partial y}[F_{X,Y}(x, y)]}{f_Y(y)}.$$

This formula is a bit messy, so let us take the partial derivative with respect to x, and this will give us the **conditional density function** of X given $Y = y$, which is

$$\begin{aligned}
f_{X|Y}(x|y) &= \frac{\partial}{\partial x} \frac{\frac{\partial}{\partial y}[F_{X,Y}(x, y)]}{f_Y(y)} \\
&= \frac{\frac{\partial^2}{\partial x \partial y}[F_{X,Y}(x, y)]}{f_Y(y)} \\
&= \frac{f_{X,Y}(x, y)}{f_Y(y)}.
\end{aligned}$$